

Extending an LIS Data Curation Curriculum to the Humanities: Selected Activities and Observations

Allen H. Renear, Molly Dolan, Kevin Trainor, Trevor Muñoz

Center for Informatics Research in Science and Scholarship

Graduate School of Library and Information Science

University of Illinois at Urbana-Champaign

501 East Daniel Street

Champaign, IL 61820

1-217-333-3280

{renear, madolan, trainor1, munoz14}@illinois.edu

ABSTRACT

We describe selected activities of an IMLS-funded project to extend an existing data curation curriculum to include humanities data, making a number of preliminary observations and conjectures.

Categories and Subject Descriptors

K.3.2 [Computing Milieux]: Computer and Information Science Education – curriculum, information systems education

General Terms

Management, Documentation, Reliability, Standardization.

Keywords

Data curation, humanities, education, curriculum, libraries

BACKGROUND

Although originally focused on data in the natural sciences, the data curation community is now concerned with data in the humanities as well. To help prepare the next generation of information professionals, the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign received a grant from the Institute of Museum and Library Services (IMLS) to extend our existing Data Curation Education Program (DCEP), which is also IMLS funded, to include humanities data. This new program, DCEP-H, began in August 2008. Among the activities underway are needs analysis studies, curriculum design, case study development, a fellowship program, an internship program, a summer institute for in-service training, and the development of a framework of cross-disciplinary data curation concepts. Through the IDEA group DCEP-H collaborates with other IMLS-funded data curation education programs to share insights and experiences. [3].

2. SELECTED ACTIVITIES

In this abstract we discuss three DCEP-H activities that are contributing to improving our understanding of humanities data curation needs. For general information about the project and about other DCEP-H activities see Renear et al.

2.1 The 2009 Summer Institute for Data Curation in the Humanities

The DCEP 2009 Summer Institute for Data Curation, held May 18-21, 2009, at the Graduate School of Library and Information Science, Urbana-Champaign, Illinois, focused on humanities data and was coordinated by DCEP-H and funded by IMLS. The Institute brought together subject specialists, archivists, digital library, metadata, and repository professionals, as well as those who work directly with research data, for a discussion of emerging themes and practices in the management and preservation of humanities data.

2.1.1 Sessions and Instructors

A number of sessions were presented, including:

- *Curation of research data: Understanding scholarly practices and the role of collections*,
Carole Palmer & Melissa Cragin (Illinois).
- *Descriptive markup and the TEI*,
Julia Flanders & Syd Bauman (Brown University).
- *Markup semantics and the preservation of intellectual content: the data curator reads Ecclesiastes*,
Michael Sperberg-McQueen (MIT/W3C).
- *Tools for Textual Data*,
John Unsworth (Illinois).
- *Digital Preservation and Institutional Repositories*,
Dorothea Salo (University of Wisconsin).
- *An appropriate institutional response?: Some organizational contexts for digital asset management*,
Lorcan Dempsey (OCLC).

Attendees

The 24 attendees of the Summer Institute came from a wide variety of institutions and roles. Of the librarians that attended seven were subject or area specialists (e.g., Special Collections, Visual Resources, Arts, English & Theatre), five were cataloguing or metadata librarians, five worked with digital content, and two were data librarians. There were three attendees from digital humanities as well as a NOAA data librarian and a content publisher.

2.1.3 What we learned

Several things stand out:

- 1) **Demand:** We received 43 applications, almost all highly qualified, and were only able to accept 24. All 24 of those accepted chose to attend.
- 2) **Satisfaction:** We were surprised at the degree of excitement and satisfaction expressed by our attendees. This is certainly another sign of unmet need.
- 3) **Interest in additional formats:** Interest was repeatedly expressed in additional in-service education, adapted to specific roles (managers, curators, technical staff), and at different levels.
- 4) **Career level:** Most of our applicants were mid-career professionals: (e.g., Data Services Librarian, Metadata Librarian, Visual Resources Librarian, Special Collections Librarian.) They were clearly preparing for anticipated management of new curatorial activities.

The demand for in-service education in humanities data curation is very high. Interest appears especially strong from managers acquiring new responsibilities, and programs focusing on those needs are in order. Valuable programs for those taking on these new roles can be developed on the model of the Summer Institute. However there is also need for a variety of formats serving different target audiences.

Towards Common Concepts

A common framework of curatorial concepts and terminology that can be applied across domains would have many advantages for data curation.

Unnecessary diversity in terminology creates a burden for design, implementation, documentation, and training. A proliferation of discipline-specific frameworks can also impede the recognition of opportunities for simplification, efficiency, and the re-use of successful strategies across domains.

Common frameworks of concepts could help (i) simplify and unify policies and documentation that support curatorial activities, and (ii) support a more uniform data curation curriculum. But can such distinct domains of scholarship as the sciences and humanities support shared frameworks of common concepts?

One important notion that appears to exist in many different domains is the distinction between data that is relatively raw or in some sense accepted as given, and data that is “processed” or the result of interpretation and analysis.

As an exercise in conceptual alignment in this area, we have compared the widely used NASA Earth Observing System data level categories with traditional notions of levels of editorial intervention found in textual criticism. The relative ease of finding intuitive alignments across domains provides some evidence that a shared framework in this area is possible. However an obstacle to a confident alignment was that neither NASA nor the literature of textual philology provides an adequate conceptual (as opposed to merely operational) account of what data levels are, focusing instead on what features should be at what level. We suspect this will be a problem for other alignment projects as well.

Given these encouraging results we are beginning a collaboration with the Data Practices and Data Concepts groups of the NSF-funded Data Conservancy. These groups, which are also based at GSLIS, are among other things developing a formal framework of cross-cutting fundamental concepts related to data curation in the sciences.

Assessing the Skills Needed for Humanities Data Curation

To determine what skills are needed by humanities data curators we have two projects that focus on the workplace:

Analysis of Position Descriptions

DCEP-H is working with an ongoing DCEP project (led by Melissa Cragin) to assess job postings in the sciences, social sciences, and humanities for skills and education relevant to data curation.[2] Job announcements were collected from a number of online job hosting sites. Relevancy was determined by the collector via a two-stage process of (i) assigning general relevancy by identifying data curation terminology and (ii) assigning specific relevancy by manually examining the announcement in detail. Analysis of the collected data will begin in January 2010.

Survey/Interview Needs Assessment

For another perspective on the skills needed for curation of humanities data we are conducting a survey and structured interviews with management and professional staff at selected humanities computing centers to identify problems faced in data management, as well as current best practices and future needs for data expertise.

ACKNOWLEDGMENTS

This program is sponsored by the Institute of Museum and Library Services (RE-05-08-0062-08).

REFERENCES

- [1] Cragin, M.H., Heidorn, P.B., Palmer, C.L., Smith, L.C. (2007). An Educational Program on Data Curation. Poster, Science and Technology Section of the annual American Library Association conference. Washington, D.C., June 25, 2007. <http://hdl.handle.net/2142/3493>.
- [2] Cragin, M., Palmer, C.L., Varvel, V., Collie, A., Dolan, M., (2009, December). Analyzing Data Curation Job Descriptions. Poster, 5th International Digital Curation Conference, London, England, December 2-4, 2009. <https://www.ideals.illinois.edu/handle/2142/14544>.
- [3] Hank, C., Davidson, J. (2009). International Data curation Education Action (IDEA) Working Group: A Report from the Second Workshop of the IDEA. *D-Lib*, 15(3/4). <http://www.dlib.org/dlib/march09/hank/03hank.html>.
- [4] Palmer, C. L. (2009, December). The Data Conservancy: A Digital Resource & Curation Virtual Organisation. Presentation, 5th International Digital Curation Conference, London, England, December 2-4, 2009.
- [5] Renear, A.H., Dolan, M., Trainor, K., Cragin, M.H. (2009). Towards a Cross-Disciplinary Notion of Data Level in Data Curation. Proceedings of the 72nd ASIS&T Annual Meeting. Vancouver, BC, November

8-11, 2009.

<http://www.ideals.illinois.edu/handle/2142/14547>.

- [6] Renear, A.H., Teffeuau. L.C., Hswe, P., Dolan, M., Palmer, C.L., Cragin, M.H., & Unsworth, J.M. (2009). Extending an LIS Data Curation Curriculum to Include Humanities Data. Poster, DigCCurr 2009 conference. Chapel Hill, NC, April 1-3, 2009.
<https://www.ideals.illinois.edu/handle/2142/14548>.